HYBRID DCA-PCA MULTIPLE FAULTS DIAGNOSIS METHOD

Funa Zhou[1,2], Tianhao Tang[2] and Chenglin Wen[3]
[1]*Computer&Information Engineering School, Henan University, Kaifeng, Henan, China*
[2]*Department of Electrical & Control Engineering, Shanghai Maritime University, Shanghai, China*
[3]*Institute of Information and Control, Hangzhou Dianzi University, Hangzhou, China*

Keywords: Unknown fault pattern, Multiple faults, DCA, PCA, Fault diagnosis.

Abstract: As it can avoid the pattern compounding problem of PCA, designated component analysis (DCA) can be used to implement multiple fault diagnosis for a multivariate process. But designated fault pattern must be defined in advance, which limited its application in unknown fault diagnosis. In this paper, a hybrid DCA-PCA method is developed for unknown multiple faults diagnosis. the main idea is: Implement DCA in the first step. Removing the designated fault pattern from the observation data, then implement PCA to the residual, and use the first loading vector as the new fault pattern to extend the fault pattern base. In the third step, implement DCA for the new fault pattern and compute the significance of the new fault pattern. Simulation for data involved 4 faults shows the efficiency of the progressive DCA fault diagnosis method.

1 INTRODUCTION

Fault diagnosis is critical for large scale system since failure in a part may cause breakdown of the system or even disastrous accident (Zhou, 2000).

In general, fault diagnosis methods can be classified into 3 classes: quantitative model-based method, qualitative model-based method and process history based method, also called data driven method (Venkat,2003, Wen, 2008, Ku, 1995).

With the widely application of DCS and intelligent instrument in industry field, it is convenient to acquire and store a large amount of data on system operation. Since these data isn't efficiently used in monitoring, it is not surprise to face "data rich, information poor" problem. People are now realizing the significance of data driven monitoring method (Venkat, 2003, Yue, 2001).

Common used data driven diagnosis method includes: expert system method, ANN based method and statistical method (Venkat, 2003, He, 2007). Among data driven methods, statistical method seems to have been well studied and applied. And PCA/PLS based methods are the dominant ones. These PCA based methods are efficient in abnormal detection. But pattern compounding effect of PCA makes it unavailable to fault pattern recognition, especially for multiple faults diagnosis (Liu, 2002).

DCA is also a multivariate statistical information extraction method. It can avoid the pattern compounding problem of PCA, thus can be used to diagnose multiple faults (Liu, 2002, Zhou, 2009).

But

1) DCA requires all the designated patterns are orthogonal, which is impractical in most application;

2) DCA diagnosis method is validated only for those known fault patterns defined in advance.

The first problem has been solved in (Zhou, 2009). This paper focuses on developing a hybrid DCA-PCA method for unknown fault diagnosis.

2 PCA BASED FAULT DIAGNOSIS

The essence of PCA is a linear transform

$$v_i = b_i^T y \tag{1}$$

Where principal component v_i is the projection of observation variable $y = [y_1, y_2, \cdots, y_p]^T \in R^{p\times 1}$ on loading vector $b_i = [b_{i1}, b_{i2}, \cdots, b_{ip}]^T \in R^{p\times 1}$, which is the ith eigenvector of y 's covariance matrix Σ_y . For a sample size of n , equa. (1) expands into the following matrix form

$$V = B^T Y \tag{2}$$

where $Y \in R^{p\times n}$ is the observation matrix, $V \in R^{p\times n}$ is the scoring matrix.

PCA decompose the observation matrix Y as a sum of p matrices of rank 1

Zhou F., Tang T. and Wen C. (2009).
HYBRID DCA-PCA MULTIPLE FAULTS DIAGNOSIS METHOD.
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics - Intelligent Control Systems and Optimization*, pages 367-370
DOI: 10.5220/0002201103670370

$$Y = \sum_{i=1}^{m} b_i v_i^T + \sum_{i=m+1}^{p} b_i v_i^T \tag{3}$$

Where m is the number of key principal component selected, $E = \sum_{i=m+1}^{p} b_i v_i^T$ is the residual.

Implement abnormal detection via the statistics T^2 and SPE (MacGregor, 1995, Zhang,2000). But they can not correctly recognize fault pattern.

3 DCA BASED FAULT DIAGNOSIS

The designated patterns are defined as $d_i = [d_{i1}, d_{i2}, \cdots, d_{ip}]^T$, where d_{ij} is 0 or 1 determined by the relation between fault and its symptom(Zhou,2009, Liu, 2004). Then, project y on d_i to get the designated components

$$w_i = d_i^T y \tag{4}$$

For a sample size of n

$$W = D^T Y \tag{5}$$

Thus Y can also be expressed as sum of p matrices of rank 1

$$Y = \sum_{i=1}^{p} d_i w_i \tag{6}$$

If there are only $l \le p$ variation pattern is designated, then (Zhou, 2009)

$$Y = \sum_{i=1}^{l} d_i w_i + E \tag{7}$$

Convergence of (7) has been proved in (Zhou, 2009).

Compute the significance of every designated pattern to determine whether the fault has occurred

$$P_i\% = \mathrm{var}(w_i) / trace(\Sigma_y) \tag{8}$$

But DCA is invalidated for unknown faults diagnosis. A hybrid DCA-PCA method will be developed to solve this problem.

4 HYBRID DCA-PCA MULTIPLE FAULT DIAGNOSIS METHOD

As it is known to all, PCA is a complete data-driven method. Although fault pattern PCA revealed makes no physical sense, it can determine a significant variation pattern of the abnormal system without any prior information. In this section, we develop a hybrid DCA-PCA multiple fault diagnosis method for the case when unknown new fault occurs.

First, define $l\ (l \le p)$ designated pattern as in (Zhou,2009), and implement DCA to the observation data; Then, remove the designated patterns defined in advance from the observation data to get the residual E ; Determine whether new fault is comprised in the residual according to the energy significance of the residual defined as $\|E\|_F$. $E = Y - \sum_{i=1}^{l} d_i w_i$ is large in the sense $\|E\|_F > \delta$ means that new fault occurred.

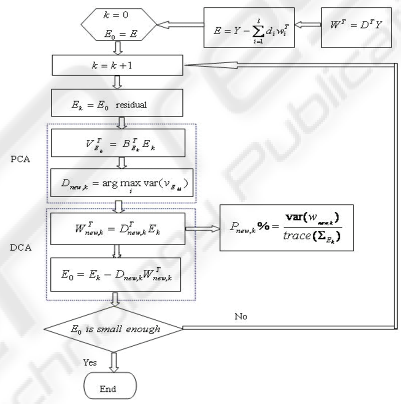

Figure 1: Hybrid DCA-PCA diagnosis method.

Implement PCA to E , and select the first loading vector as new fault pattern $D_{new,1}$. Another round of DCA is carried out for $D_{new,1}$. Repeating this process until residual is insignificant.

$$W^T = D^T Y \tag{9}$$

$$E = Y - \sum_{i=1}^{l} d_i w_i \tag{10}$$

$$V_E = B_E^T E \tag{11}$$

$$D_{new,1} = \max_i \mathrm{var}(v_{E_i}) \tag{12}$$

$$E_1 = E \tag{13}$$

$$W_{new,1} = D_{new1,1}^T E_1 \tag{14}$$

$$E_2 = E_1 - D_{new,1} W_{new,1} \tag{15}$$

$$V_{E_2} = B_{E_2}^T E_2 \tag{16}$$

$$D_{new,2} = \max_i \mathrm{var}(v_{E2_i}) \tag{17}$$

$$\vdots$$

Figure 1 depicts the hybrid DCA-PCA multiple fault diagnosis process.

5 SIMULATIONS

Simulations parameters used are: $p = 15$, $n = 1000$, $l = 6$.

The observation data is generated by the composition of 10 variation patterns

$$Y = \sum_{i=1}^{10} d_i \overline{w}_i \tag{18}$$

Where $\overline{w}_i \sim N(0, \sigma_i^2)$ is the simulated designated component, $d_i \in R^{p\times 1}$ $(i = 1,2,\cdots,10)$ are the 10 variation patterns to generate observation data Y. figure 2 depicts the contribution of each d_i to Y.

The first 6 pattern $d_1,\cdots,d_6$ are the designated pattern we selected, d_1, d_3, d_5 is the fault pattern, d_{10} is a fault pattern unconsidered in advance.

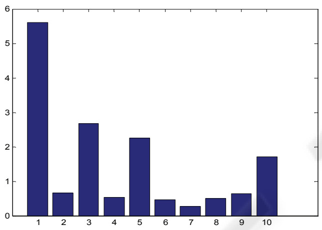

Figure 2: DC values for Generating observation.

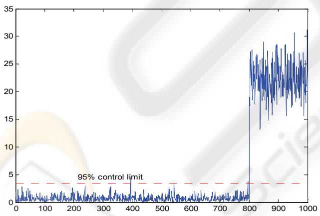

Figure 3: SPE chart for PCA.

For normal observation, $\overline{w}_i$ can be generated in MATLAB using "randn" and some linear operation

For fault case, d_1, d_3, d_5, d_{10} are amplified from sampling time 801 to 1000

$$\overline{w}_i = \overline{w}_i + 5\,\mathrm{var}(\overline{w}_i) \quad i = 1,3,5,10 \tag{19}$$

5.1 PCA based Fault Diagnosis

The SPE chart of PCA are shown in figure 3. Figure 3 indicates that system considered is abnormal from 801st sample point. However, SPE chart can not tell what faults occur.

5.2 Hybrid DCA-PCA Multiple Faults Diagnosis

Implement DCA to the observation, and illustrate the significance of each designated pattern in table 2. From table 2, we can see that, the fist, the third and the fifth variation pattern are the 3 significant ones of the 6 designated patterns. According to the significance of the observation to each designated, we can conclude that faults corresponding to these 3 designated patterns have occurred in the system.

Table 2: Significance of the designated pattern in D.

	d_1	d_2	d_3	d_4	d_5	d_6
$d_i\%$	0.41 23	0.036 8	0.1946	0.035 9	0.1839	0.037 8

Figure 6 draws the Shewhart chart of every designated component in. It indicates that the 1st, 3rd, and 5th designated component's Shewhart chart exceed the control limit from 801 to 1000. Figure 6, confirms that faults corresponding to the 1st, 3rd, and 5th designated patterns have occurred in the system.

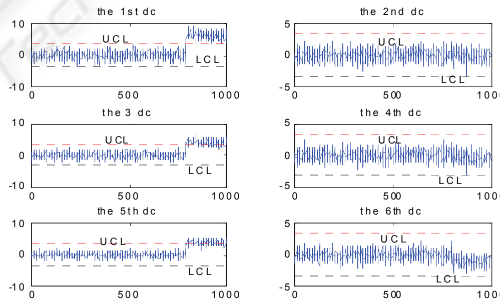

Figure 6: Shewhart chart for 6 DC.

For the case $p = 15$, $n = 1000$ and $l = 6$, statistical result of more than 100 times simulation shows that the threshold is reasonable

$$\delta = 10 \tag{20}$$

Removing the 6 designated variation pattern to get the residual E_0. The norm of the residual is

$$\|E_0\|_F = 12.7524 > \delta \tag{21}$$

It is possible that at least one new fault is still included in the residual. Implement PCA to the residual E_0, and take the first load vector as a new fault pattern. Then implement DCA to E_0 for d_{new1}. And compute its significance

$$d_{new1}\% = 0.6083 \quad (22)$$

The Shewhart chart of this new designated component is depicted in figure 7. Figure 7 tells us that fault corresponding to d_{new1} has occurred in the system. Removing the new fault pattern d_{new1} from E_0 we have the residual of this DCA step

$$\|E_1\|_F = 7.2042 \quad (23)$$

To the residual E_1, Shewhart chart for the secondnew designated component, figure 8 is within the control limit, which will confirm that $\delta = 10$ is reasonable

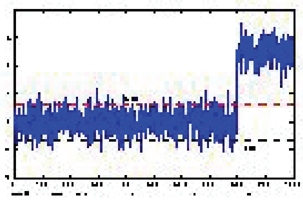

Figure 7: Shewhart chart for the 1st new dc.

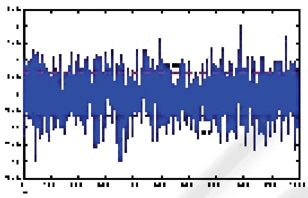

Figure 8: Shewhart chart for the 2nd new dc.

From the above simulation research, we can conclude that d_1, d_3, d_5 and d_{10} occurred in the system. This is basically the same as the simulation manner that we used to generate Y.

6 CONCLUSIONS

DCA can avoid pattern compounding problem of PCA. But it is invalidated for unknown faults diagnosis. In this paper, a hybrid DCA-PCA method for unknown multiple fault diagnosis.

Some data driven methods other than PCA can be used to the residual to estimate the new fault pattern to make it physical sense.

ACKNOWLEDGEMENTS

This paper is supported by NSFC (60804026); International cooperation project of Zhejiang (2006C24G2040012), Natural science fund of Henan (2009A510001) International cooperation project of Henan (094300510043), Key disciplines of Shanghai Municipality (J50602), Development Project (08YZ109) from Shanghai Municipal Education Commission.

REFERENCES

Donghua Zhou, Yinzhong Ye, 2000. Modern fault diagnosis and tolerant control[M], Beijing, Qstinghua Publishing House (in Chinese).

Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewn Yin, Surya N. Kavuri,2003. A review of process fault detection and diagnosis Part I[J]: quantitative model-based methods. Computers and Chemical Engineering 27 (2003):pp293-311.

Qingbo He, 2007. Application Multivariate statistical analysis in machine state monitoring and diagnosis [D], PHD thesis, University of Science and Technology of China (in Chinese).

Yue H H, Qin S J, 2001. Reconstruction based fault identification using a combined index. Industrial and Engineering Chemistry Research[J], 40(20): 4403-4414

J.F. MacGregor and T. Kourtl, 1995. Statistical process control of multivariate processes[J], Control Fag. Practice, VoL 3, No. 3, pp. 403-414.

Yegang Liu, 2002. Statistical control of multivariate processes with applications to automobile body assembly (D). PHD, University of Michigan.

Jie Zhang, Xianhui Yang. Multivariate statistical control[M], Beijing, Chemistry Industry Publishing House, 2000 (in Chinese).

Funa Zhou, Chenglin Wen, Tianhao Tang, 2009. DCA based multiple faults diagnosis method, accepted by ACTA AUTOMATICA SINICA (in Chinese).

Chenglin Wen, Jing Hu, Tianzhen Wang, Zhiguo Chen. RPCA and it's application in data compression and fault diagnosis, ACTA AUTOMATICA SINICA 34(9) : 1128-1139 (in Chinese).

Ku, W., Storer, R.H., and Georgakis,C, 1995. Disturbance detection and isolation by dynamic principal component analysis [J], Chemometrics and Intelligent Laboratory Systems, 30:179-196.